STIF

01

"Made available under NASA sponsorship in the interest of early and wide dissemination of Earth Resources Survey Program information and without liability for any use made thereof."

II

ERIM

FORMERLY WILLOW RUN LABORATORIES, THE UNIVERSITY OF MICHIGAN

7.6-10502
CR-148822

WHEAT PRODUCTIVITY ESTIMATES USING LANDSAT DATA

TYPE II PROGRESS REPORT

16 May 1976 - 15 August 1976

114800-20-L

NASA Contract No. NAS5-22389

Prepared by

Richard F. Nalepka - Principal Investigator

John Colwell - Co-Principal Investigator

Daniel P. Rice

(E76-10502) WHEAT PRODUCTIVITY ESTIMATES USING LANDSAT DATA Progress Report, 16 May - 15 Aug. 1976 (Environmental Research Inst. of Michigan) 23 p HC $3.50 CSCL 02C

N76-32615

Unclas

G3/43 00502

for

Mr. G. R. Stonesifer, NASA Technical Officer
Code 902
National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt Road
Greenbelt, Maryland 20771

2062L

WHEAT PRODUCTIVITY ESTIMATES USING LANDSAT DATA
TYPE II PROGRESS REPORT
16 May 1976 - 15 August 1976

The following report serves as the fifth Type II Progress Report for Landsat Follow-on Investigation #2062L which is entitled "Wheat Productivity Estimates Using Landsat Data."

This investigation has several objectives, including the following:

1) to develop techniques and procedures for using Landsat data to estimate characteristics of wheat canopies which are correlated with potential wheat grain yield.

2) to demonstrate the usefulness of Landsat data for estimation of wheat yield

a) for irrigated and for non-irrigated LACIE (Large Area Crop Inventory Experiment) intensive test sites.

b) for two different years with varying weather conditions.

A. PROBLEMS

None.

B. ACCOMPLISHMENTS AND RESULTS

On the following pages we discuss the many technical areas addressed during the reporting period.

Field Work

Field data collection efforts for Finney County Kansas have continued with a mission centered around the June 2 Landsat overpass. Photographic records for determination of percent vegetation cover were obtained on 12 fields for which actual yield is to be determined. Samples of wheat leaves were harvested and taken to ERIM for measurement of their radiometric properties (reflectance and transmittance) on a Beckman spectrophotometer. Surface soil samples were also collected on

several fields and were returned to ERIM for measurement of spectral reflectance on the Cary 14 spectrophotometer.

As of this writing the soil reflectance measurements have been made, and have been used to help guide the processing of the Landsat data. The leaf radiometric properties have been measured, but have not yet been reduced to hemispherical reflectance and transmittance values. Some of the field photos have been reduced to percent cover values.

Ancillary Data

Ancillary environmental data have been obtained from both the Finney and Ellis County Kansas sites. This data includes information such as maximum and minimum daily temperature, and will help to correct for differences in timing of phenological events from one place to another (e.g., Finney → Ellis) and from one time to another (e.g., 1975 → 1976).

Data Handling

During this reporting period, the field mean signal values in each Landsat band were extracted for all sufficiently large fields for 5 Ellis County Kansas scenes, and stored in a data base. The variables stored in the data base for each field include:

a. ground truth parameters such as crop condition, yield, etc.
b. number of pixels extracted from the field, in each time period
c. the Landsat channel mean signal values in each time period computed from the pixels extracted from the field
d. corresponding EXTEC3-transformed data (see Appendix A)
e. a "green development" and a "soil-brightness" feature mean, also computed by EXTEC3.

By storing the data in this form we have greatly increased the ease of statistically analyzing the data and developing methods of predicting yield.

Data Normalization

In order to estimate yield from one data set using a relationship established on a different date or place, the separate Landsat data

sets must be "normalized" to equivalent values by removing any non-target related effects such as those due to the atmosphere, solar irradiance, and the like. Two methods of data normalization were investigated during this reporting period. One method is based on a matching of data patterns by hand, and the other is the EXTEC3 procedure described in Appendix A. Both methods are briefly discussed in this section.

The first method of normalization that was tested is carried out by a visual inspection of two channel scatter plots of the two data sets to be normalized. Figure 1 is an example of these scatter plots, showing the Landsat Band 5 versus Band 6 pattern. By comparing the pattern of the 20 May Ellis data set to that of the 21 May Ellis data set in each pair of adjacent channels, one can determine approximately how much relative displacement exists, on the average, in each channel. These displacements were subtracted from data points in one scene to normalize that scene to the other. While this method is subject to variability of human judgement, an initial test showed reasonable agreement when different persons independently determined the displacement. More sophisticated corrections of this type are also being investigated.

A second normalization method that was tested, EXTEC3 (which was developed using Landsat 1 data), was applied to five 1975 Landsat scenes of the Ellis County Kansas test site for the dates 3 May, 11 May, 20 May, 21 May, and 17 June, and the results were assessed by comparing two channel scatter plots of the EXTEC3-transformed data. It was found for the 20 May and 21 May scenes that the overall pattern of pixels on the Band 5 versus Band 6 transformed data showed a remaining displacement between the patterns noticeably less than that present in the untransformed data. Even a greater improvement can be expected by further adjusting the parameters of EXTEC3 to optimize performance. An additional need with respect to EXTEC3 is to determine parameters

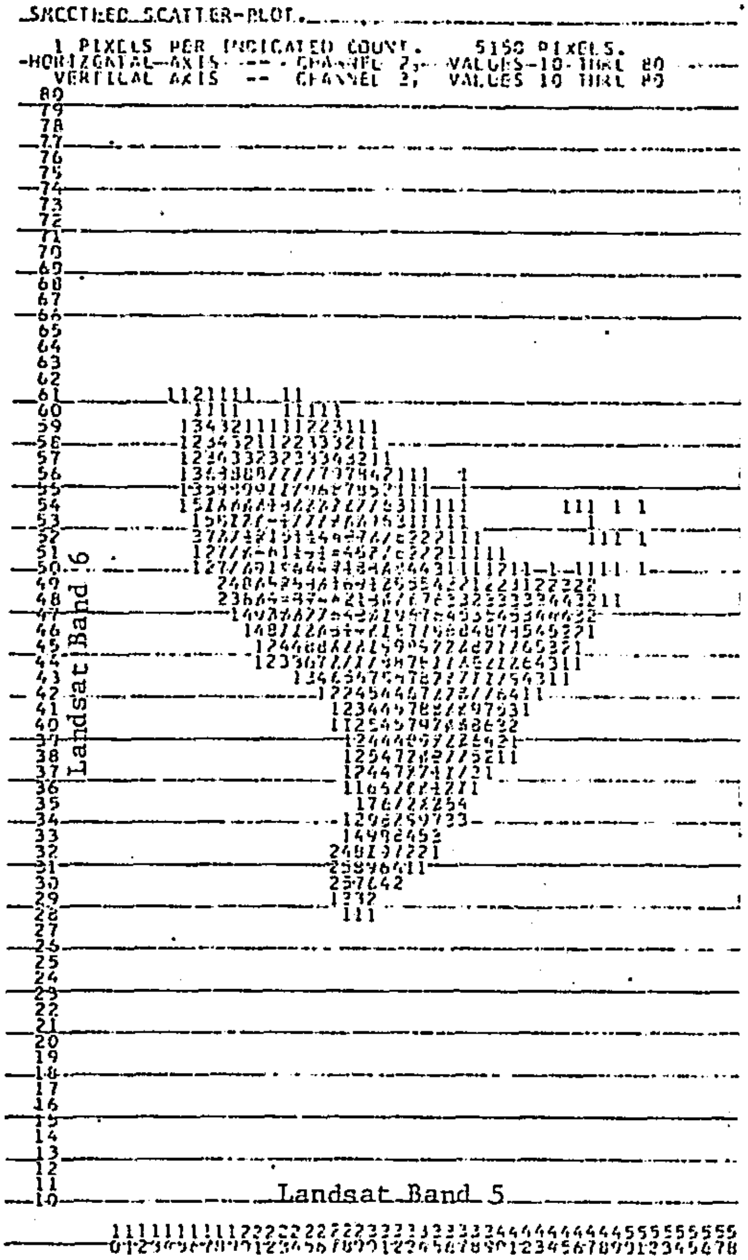

FIGURE 1. EXAMPLE SCATTER PLOT USED IN DATA NORMALIZATION - ELLIS COUNTY, KANSAS, MAY 21, 1975 TEST SITE; ALL PIXEL USED. (The frequency of occurrence of pixels at each position is specified by the symbol used)

appropriate for Landsat 2 data, since there are significant scaling differences between the data from Landsat 1 and Landsat 2 which must be compensated.

We now turn to the matter of testing the usefulness of the normalization procedures.

First, the effect was examined of not normalizing the data. This was accomplished using May 20 and May 21 Landsat data sets of the Ellis site. Adjacent day data was chosen since it was felt that this would probably minimize normalization problems, thereby providing a base value for the severity of the problems. It is reasonable to assume that crop development in the test wheat fields changed little during the two adjacent days while atmospheric conditions were somewhat different and the look angle was different. The test of the need for normalization consisted of determining the utility of a relation for predicting yield on May 20 Landsat data which was developed on May 21 Landsat data.

The best performance that could be expected in predicting yield using May 20 Landsat data was determined by a linear least squares regression of yield vs the four May 20 Landsat data channels. The mean square error* (MSE) for this regression using 24 fields was calculated by

$$MSE = \frac{1}{n-m-1} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

* The MSE is one commonly computed statistic for assessing the "goodness" of a regression in terms of the difference between actual and predicted values. Simple correlation statistics are not sufficient for this analysis since they remain unchanged when a linear transformation such as the hand-normalization method is applied to the data.

where

n = number of cases (fields) [=24]

m = number of variables (channels used in regression) [=4]

Y_i = yield for field i

$\hat{Y}_i$ = predicted yield for field i

The base MSE that resulted for the May 20 Landsat data was 29.0. A similar regression was then performed using the May 21 Landsat data. The resulting regression equation was then applied unchanged to the May 20 Landsat data to predict yield, and the mean square error was again calculated. This MSE value was found to be 149.5. Clearly, much of the predictive capability was lost when the data sets were not mutually normalized.

The May 20 Landsat data was subsequently manually normalized to the May 21 data by subtracting the amount of apparent relative displacement from the May 20 field means, after examining the scatter plots. The regression equation determined from the May 21 data was then applied to the hand-normalized May 20 data, and the MSE value was calculated again. The resulting value of 39.8 in this case is only slightly larger than the base May 20 result of 29.0. For a comparison of the MSE values, one may refer to Table 1.

In order to statistically quantify the degree to which performance is degraded in extending a yield predicting regression equation from one data set to another, an "F-statistic" was computed as the ratio of MSE of the extended equation to the base equation (Table 1). The larger the F-ratio, the worse the prediction extension performance is compared to the base prediction performance. In a statistical sense, the reference and extended data sets (and hence the required regression equations) are assumed different at the 5% level of significance if $F>2.17$, and are assumed different at the 1% level if $F>3.03$ (for 19

TABLE 1. RESULTS FOR TESTS OF THE DATA NORMALIZATION PROCEDURES (Ellis County, Kansas, Test Site; 24 Wheat Fields)

	DATA ON WHICH YIELD RELATION WAS DERIVED	DATA ON WHICH YIELD RELATION WAS APPLIED	MSE IN PREDICTION OF YIELD $\sum(Y-\hat{Y})^2/n-5$	F-STATISTIC (Ratio of MSE Relative to Base (May 20) MSE)
1	May 20 Unnormalized	May 20 Unnormalized	29.0	(Base)
2	May 21 Unnormalized	May 20 Unnormalized	149.5	5.16**
3	May 21 Unnormalized	May 20 Hand-normalized	39.8	1.37
4	May 20 EXTEC3	May 20 EXTEC3	33.2	1.14
5	May 21 EXTEC3	May 20 EXTEC3	37.5	1.29

**Exceeds the threshold for the 1% significance level of 3.03 for 19 degrees of freedom.

degrees of freedom). Since the F-statistic for predicting yield from unnormalized May 20 Landsat data using the May 21 regression equation exceeds both significance thresholds, the two sites are considered too different for effective yield prediction extension without data normalization. In this trial, however, the hand-normalization appears to be effective, since the F-statistic is much less than the threshold.

With a second normalization technique, EXTEC3, both May 20 and May 21 data were normalized to a standard (hypothetical) Landsat data set, and hence, were normalized with respect to each other. The linear regression of yield versus the four May 20 EXTEC3 data channels was computed and was found to have a MSE of 33.2. This result is slightly poorer than using the original data, as there has apparently been some loss of information in the EXTEC3 transformation process. A linear regression was then performed on May 21 EXTEC3-transformed Landsat data, and the resulting regression equation was applied to the May 20 EXTEC3-transformed data. The MSE in predicting yield using the May 20 EXTEC3-transformed data was then found to be 37.5. At first glance, it might appear as though the EXTEC3 normalization procedure exceeds hand-normalization in performance (Table 1). However, high and low values of yield were predicted less accurately using EXTEC3 for this particular data set. The significance and generality of this behavior are still being investigated.

The results of the above discussion are presented in Table 1, from which it is clear that some form of normalization of the data is required to obtain improved results.

Feature Enhancement

Previous experience has suggested that individual Landsat spectral bands could have quite different values for identical values of vegetative cover and potential yield, and that one of the most important causes of this ambiguity was variation in soil spectral reflectance. Such a situation is clearly undesirable, since it prevents a unique

association of vegetative condition and Landsat data values.

One way that has been suggested to alleviate this problem is to form a ratio of an infrared and a red channel, which in many situations tends to reduce variations due to varying soil reflectance. The ratio also retains much of the information regarding the vegetative development (percent cover, LAI*) of the wheat canopy, and may even help to normalize data with respect to such factors as variations in solar irradiance, ground slope, and the like.

In order to determine whether an infrared/red ratio would be effective on Kansas soils, we collected samples and made spectral reflectance measurements of a variety of soils from both the old (1975) and new (1976) Finney Intensive Test Sites. The results for the 1976 data (Table 2) suggest that ratio processing can be effective in normalizing variations in soil reflectance for soil conditions found in Finney County, Kansas. The reflectance ratio of wavelengths 0.75 μm/0.65 μm (approximately equated to Landsat Band 6/Band 5) seems to be the best in this respect. However, preliminary analysis suggests that Landsat Band 7 is better than Band 6 as an indicator of vegetative development and potential yield, presumably due to the greater contrast between vegetation and soil in Band 7. Therefore, a Band 7/Band 5 ratio may be more useful for simultaneously reducing significant soil reflectance variation and enhancing for differences in vegetative development. Both 7/5 and 6/5 ratios are being tested using Landsat data to predict wheat yield. Initial analysis of their relative usefulness has produced results which are not conclusive.

Another transformation of the Landsat data which is being tested for its yield/vegetative development prediction capabilities is computed as part of the EXTEC3 program. EXTEC3 generates two hybrid axes (directions), including one that is nominally in the direction of green development, and another in the direction of variation in soil-

*Leaf Area Index

TABLE 2. AVERAGE SOIL SPECTRAL REFLECTANCES AND REFLECTANCE RATIO (m), AND CORRESPONDING COEFFICIENTS OF VARIATION (σ/m), FOR 19 SOIL SAMPLES TAKEN FROM THE NEW FINNEY SITE

Wavelength (nm)

650		750		900		750/650		900/650	
m	σ/m	m	σ/m	m	σ/m	m	σ/m	m	σ/m
20.75	0.53	24.81	0.49	29.18	0.41	1.24	.09	1.53	0.16

brightness. The soil brightness channel is approximately orthogonal to the "green development" channel. If the green-development channel adequately defines the extent of vegetative development, it should provide a valuable indication of potential yield. Furthermore, it is a direction that in theory can be uniquely and consistently defined for all Landsat data sets.

Initial testing of the information content in the green development channel suggests that the single direction may not be completely satisfactory for quantifying degree of vegetative development or yield. In fact, there seems to be a considerable amount of yield-predicting information in the soil-brightness channel, which is a measure of overall scene brightness. This situation may be due to an increase of shadowing within the canopy as the amount of green vegetation increases, which tends to decrease the overall scene brightness. In addition, there is possibly a correlation between soil reflectance and vegetative development and yield. In non-irrigated areas, the brighter soils may be the sandier soils, with less available stored water and with less available nutrients. The darker soils may contain more clay and so hold more moisture and possible nutrients. However, it may be risky to take advantage of this information, because other conditions can affect soil brightness but have opposite correlation with yield, and because undetectable soil conditions (e.g., fertilization, subsurface moisture) can cause differences in growth but not in soil brightness.

The relative usefulness of the green-development and soil-brightness channels, and of the Band 7/Band 5 and Band 6/Band 5 ratios, as well as other possible features, are being examined for their ability to account for yield on a particular data set and also for predicting yield using the same equation on a different data set.

Temporal Analysis (Ellis 1975)

Landsat data, even if not normalized, can be analyzed for relative information content in predicting yield. Since the spectral-temporal

information content of Landsat data for predicting yield is of considerable interest, that topic will be addressed.

The 20 individual spectral-temporal Landsat bands from five 1975 Ellis scenes (May 3, May 11, May 20, May 21, June 17) were correlated with each other and with farmers' estimates of wheat grain yield. The correlations with yield as a function of time are indicated in Figure 2. The horizontal dotted lines are 5% significance lines, so that correlation values which fall between the dotted lines are not considered significant at the 5% level. The single best spectral-temporal band for predicting yield is a May 20 red band (Band 5, 0.6-0.7 μm), with the May 21 red band a close second. Each of the visible (green or red) spectral-temporal bands is significantly correlated with yield. Fewer of the infrared bands (Bands 6 and 7) are significantly correlated with yield, and the correlation changes from positive to negative during the period of time from May 21 to June 17. This latter fact may be due to senescence of leaves over this period of time. On June 17 primarily vertical components of the canopy stalks and heads remain and a greater density of such vertical components could result in more shadow and a darker canopy. This may be the cause of the negative correlation between the near-IR bands on June 17 and harvested grain yield (which is correlated with number of stalks).

The optimum combination of spectral-temporal bands for predicting yield was determined by stepwise regression. Although the red bands (Band 5) on May 20 and May 21 are the two best individual bands, the best combination of two bands is the May 20 red band the May 11 Band 7 (0.8-1.1 μm). These two bands are negatively correlated with each other (-0.60) and together they account for 68% of the variance in yield (coefficient of variation = R^2), using a linear regression.

All four Landsat spectral bands from each of the five different dates were regressed against yield in order to assess the single best date for predicting wheat grain yield using all four bands. The results

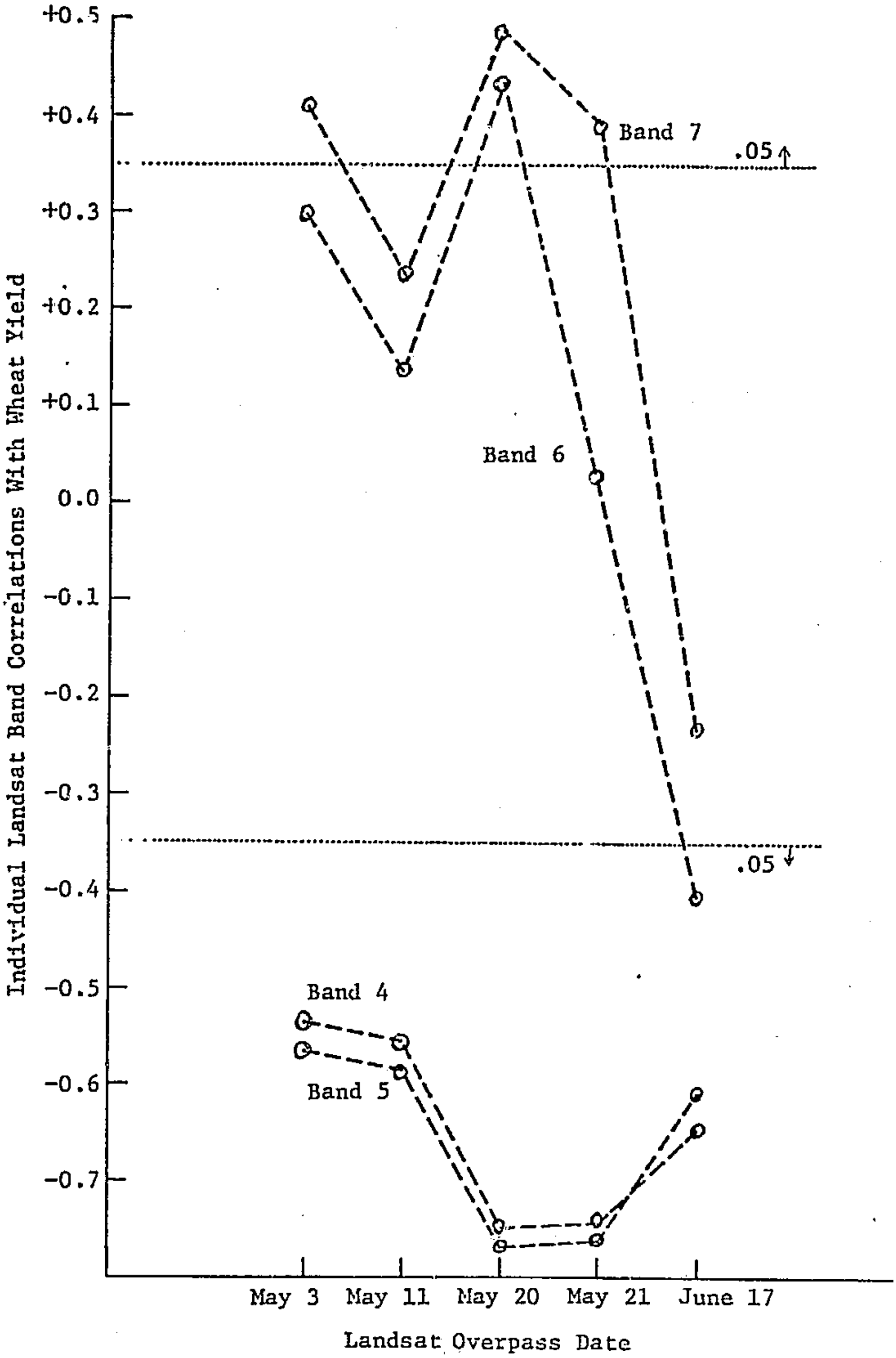

FIGURE 2. CORRELATION OF INDIVIDUAL LANDSAT BAND WITH WHEAT YIELD FOR 5 DATES. An average over 33 fields of two pixels or more with a pixel inset of 1.0 was used for each Landsat band for each date. The horizontal dotted lines specify the 5% significance level (Ellis County Kansas site).

are presented in Figure 3. The best single date is May 21, which is near, but slightly before the time at which most of the fields are in the heading stage. Not surprisingly, May 20 is a close second for choice of optimum date. The utility of the four spectral bands on the optimum single date (May 21) for predicting yield was then compared to that of the best four spectral-temporal bands. The four spectral-temporal bands were judged to be better, since the four spectral bands from May 21 account for about 69% of the variance in yield, compared to 74% for the optimum four spectral-temporal bands. The 15 best spectral-temporal bands of those investigated account for over 90% of the variance in yield using a linear least squares regression (see Figure 4). In other words, most of the variance in yield can be accounted for by Landsat data covering the early May to mid-June time span.

The foregoing analysis suggests that temporal Landsat data is important for predicting wheat grain yield. It also suggests that data near the point of heading is more useful for predicting wheat grain yield than data earlier or later in the year. The May 3 data set appears to be the least useful single date of those studied for predicting yield, accounting for only 36% of variance in yield as opposed to the 69% on May 21. The above evidence suggests that the timing of the Landsat data collection is rather important.

Selecting Fields and Pixels for Analysis

In order to form valid Landsat signal mean values for each field, we must determine which pixels are to represent that field. We must avoid using any pixels which are so near the boundary of a field as to risk containing any signal from the boundary or adjacent field. And yet we wish to select a sufficient number of fields, with a sufficient number of pixels within each field and sufficient range of yield values, so as to carry out meaningful analyses. Unfortunately, when data are so limited, a compromise between the above desires is required. The discussion which follows describes our efforts to achieve the best compromise.

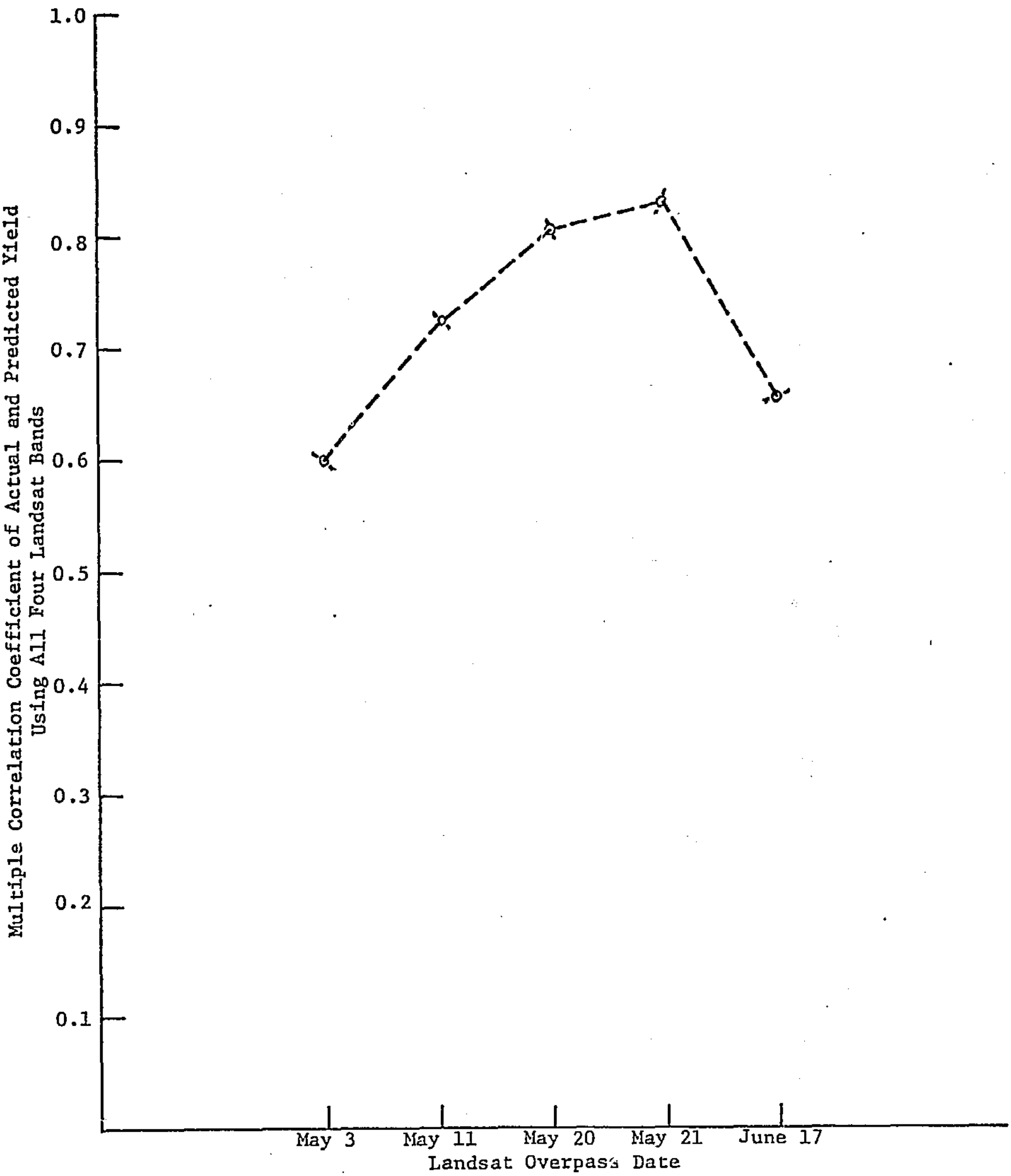

FIGURE 3. MULTIPLE CORRELATION COEFFICIENTS BETWEEN PREDICTED AND ACTUAL YIELD USING THE SET OF 4 LANDSAT BANDS FOR EACH OF 5 DATES. An average of 33 fields of two pixels or more with a pixel inset of 1.0 was used for each date. (Ellis County, Kansas site)

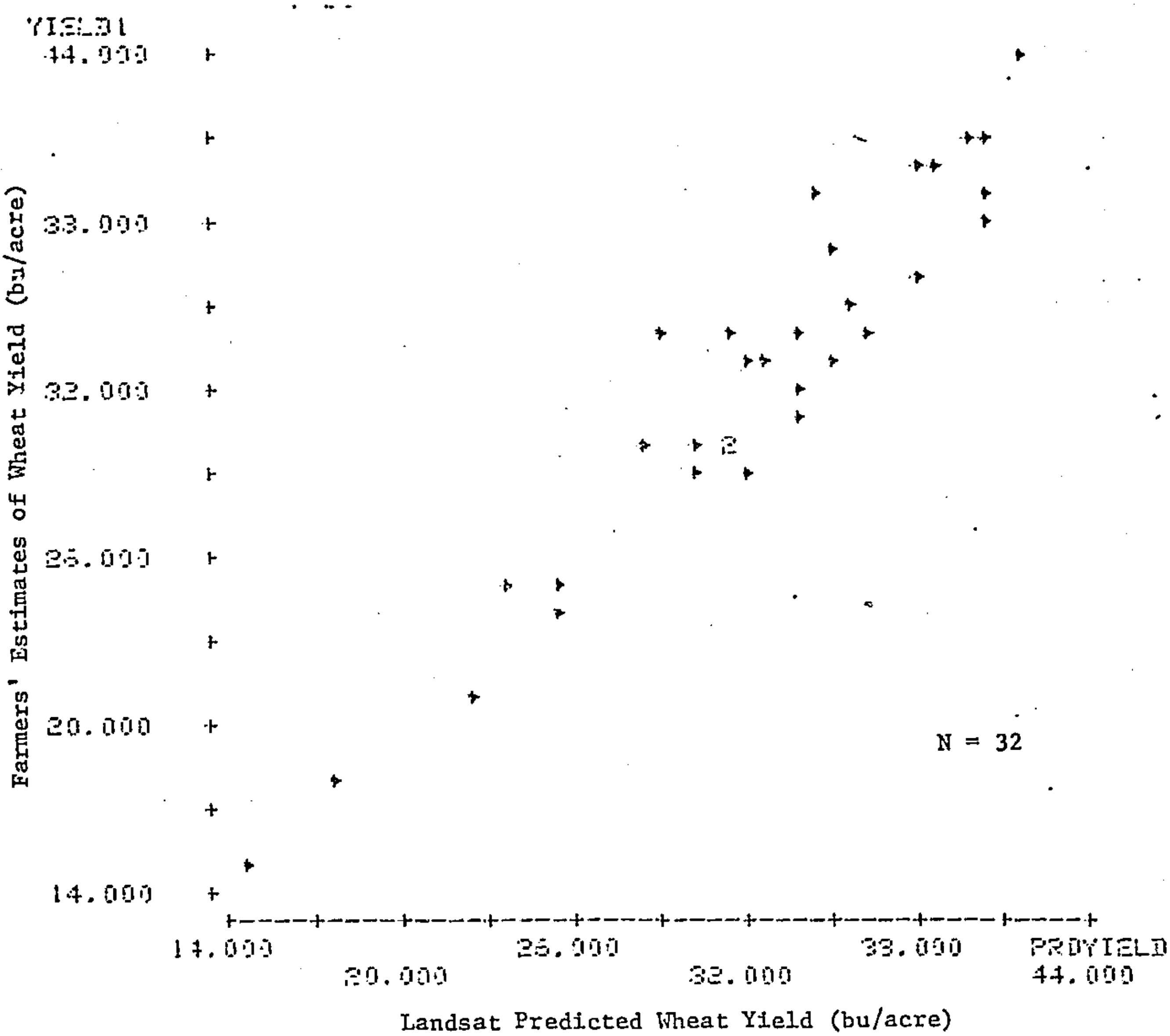

FIGURE 4. SCATTER PLOT OF ACTUAL WHEAT YIELD VS PREDICTION OF WHEAT YIELD USING THE OPTIMUM 15 SPECTRAL-TEMPORAL BANDS. (Ellis County, Kansas site)

For much of our analysis so far with the Ellis Landsat data, we have used pixel inset distance of 1.5 pixel diameters*, which means that the center of a pixel considered safely within the field must be at least 1.5 pixel diameters within the nearest edge of the field. This guarantees a one pixel separation between the pixel edge and the field edge to guard against error in the location of the field boundary, and therefore in using boundary pixels. This very conservative distance would frequently be used when pixels are relatively plentiful, or when field location errors are believed to be as much as one pixel.

In the case of our data, we believe the field boundaries are located to an accuracy usually better than 0.5 pixels. Therefore, we can with reasonable safety use an inset distance of 1.0 pixels. By so doing, we have increased the number of fields that have at least one pixel, from 24 (when inset of 1.5 was used) to 36 (with the 1.0 inset). In addition, we have thereby included fields with yield less than the previous minimum of 24.5 bu./acre, so that now the available range of yield values starts at 15.0 bu./acre, an increase of approximately 50% in the range of yield values represented.

The standard deviations of the field mean values computed with 1.0 and 1.5 pixel insets were not appreciably different. The mean values varied by an average of less than ± 0.5 digital counts. Thus, we suffered no serious deficiency by using a 1.0 pixel inset, but have received significant advantage.

An additional consideration was to decide on a rule for accepting fields, based on the number of pixels selected from each field. Unfortunately, we discovered a positive correlation between number of pixels per field and field yield. In order to retain information for the fields with the lowest yields, it was necessary to accept any field

*A pixel diameter is the distance between two adjacent pixels in a scan line, or the distance between two adjacent scan lines, using an aspect ratio for which the two distances are equal.

with no fewer than two pixels for every date. Keeping a broad range of yield values is considered sufficiently important that for most analyses, a two pixel criterion was chosen as the preferred compromise. The criterion resulted in the elimination of four of the 36 fields from further analysis. Any more stringent requirement for number of pixels would have increased the lowest value of yield in fields to be accepted to 21.4, not much below the value for a 1.5 pixel inset.

C. FUTURE PLANS

A high priority for the immediate future is the verification of a consistently effective data normalization procedure. Adequate data normalization is essential for extrapolation of a yield prediction relationship over time and space. Once an improved data normalization procedure is demonstrated, a test of the generality of a Landsat yield algorithm will consist of an attempt to predict yield on 1975 Finney data by applying a relationship developed for 1975 Ellis data.

Reduction of field data collected during the 1976 growing season will continue. Processing of 1976 data for the Finney site will begin soon after the data currently on order arrives.

D. FUNDS EXPENDED

Total expenditures during the period 16 May 1976 through 15 August 1976 are $25,307.

E. DATA USE

The following table represents the status as of 15 August 1976.

	Value of Data Allowed	Value of Data Ordered	Value of Data Received
USDI EROS Data Center	$18,000	$6,400	$4,000
USDA/ASCS Aerial Photography Field Office	$ 4,000	$1,323	$1,003

APPENDIX A
THE EXTEC3 ALGORITHM

A technique called EXTEC3 has been developed jointly by this project and others* to correct Landsat scenes for the effects of variable haze. The objective is to force data in each scene to match a standard scene, so that in all scenes a specific reflectance of the target results in a specific Landsat data value. Fulfillment of this objective would reduce the error, due to haze differences, of estimating parameters (such as vegetative ground cover) from the data.

The basis of the technique is that the four-channel data lies primarily in a single two-dimensional plane in signal space, and that the position of that plane shifts, and the pattern of pixels on the plane shrinks, as haze level is increased. The effect is approximated by specifying a reference plane (which is the two-dimensional plane on which the pixels of a "standard" data set lie), and specifying a "point of haze" toward which data would shift and shrink if made more and more hazy. Then, as shown in Figure 5, the data is projected onto the reference plane by rays extending from the point of haze.

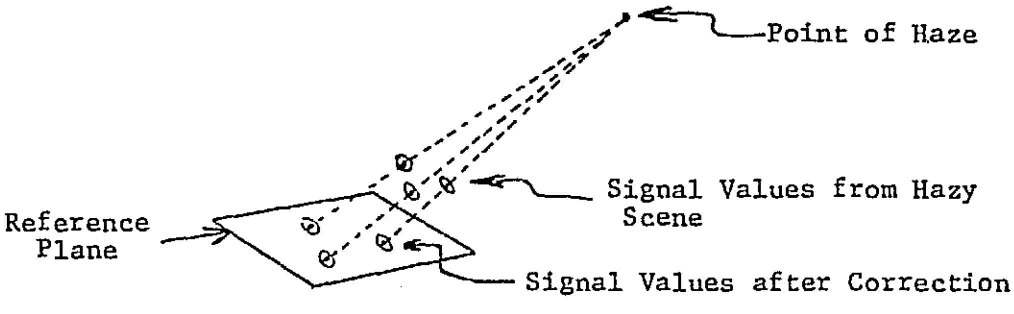

FIGURE 5. EXTEC3 METHOD OF HAZE CORRECTION

*A part of this development is being supported on NASA Contract NAS9-14988 with NASA/JSC.

The mathematics required to perform the indicated transformation is as follows.

Let:

x_h = signal value of the point of haze.

x_o = signal value of some point on reference plane.

v_h = unit vector normal to reference plane, parallel to a perpendicular dropped from x_h to the reference plane.

x = signal value of a pixel in the scene to be transformed.

y = signal value of the pixel after transformation.

The transformation is:

$$y = \frac{v_h^T(x_o - x_h)}{v_h^T(x - x_h)} (x - x_h) + x_h$$

The values used for x_h, x_o, and v_h used in the initial test are:

$$x_h = \begin{bmatrix} 89.9 \\ 71.6 \\ 61.4 \\ 23.2 \end{bmatrix} \qquad x_o = \begin{bmatrix} 48.5 \\ 51.5 \\ 53.9 \\ 24.8 \end{bmatrix} \qquad v_h = \begin{bmatrix} -.85 \\ .51 \\ .05 \\ .06 \end{bmatrix}$$

As a part of EXTEC3, two features are computed for each pixel -- a "soil brightness" and a "green-stuff" feature. Soil brightness b is measured in the direction of typically greatest soil variability, as computed by:

$$b = R_1^T y + k$$

where

$$R_1^T = (.433 \quad .632 \quad .586 \quad .264) \quad \text{[soil-brightness direction]}$$

and

$$k = \text{scaling constant} = 200 - R_1^T x_o.$$

Green-stuff is meant to represent the amount of green vegetative development, and is measured in the reference plane approximately perpendicular to the soil direction R_1. The computation is:

$$s = 32 + R_2^T \quad w$$

where

$$w = x_{SH} + \frac{200}{k} (y - x_{SH})$$

and

$$x_{SH} = x_o - 200 \ R_1$$

and

$$R_2^T = (-.289 \quad -.562 \quad .599 \quad .491) \quad \text{["green vegetation" direction]}$$

Submitted by: Richard F. Nalepka
Richard F. Nalepka
Principal Investigator
Head, Multispectral Analysis Section

Approved by: Quentin A. Holmes
Quentin A. Holmes
Head, Information Systems and Analysis Department

Approved by: Paul Rice for
Richard R. Legault
Director, Infrared and Optics Division